

See corresponding article on page 1379.

## Food substitutions revisited

Daniel B Ibsen<sup>1,2,3,4</sup> and Christina C Dahm<sup>3</sup>

<sup>1</sup>Medical Research Council Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom; <sup>2</sup>Steno Diabetes Center Aarhus, Aarhus, Denmark; <sup>3</sup>Department of Public Health, Aarhus University, Aarhus, Denmark; and <sup>4</sup>Department of Nutrition, Sports and Exercise, University of Copenhagen, Frederiksberg, Denmark

The concept of food substitutions is simple: eating 1 food instead of another. Whereas studying the effects of eating 1 food instead of another is typically explicit in interventional study designs, it is often implicit and sometimes hidden in analyses of observational studies. The substitution in question is often introduced as a consequence of adjusting regression models assessing associations between a food and an outcome of interest for total energy intake, thus restraining the comparison within the model to participants reporting the same energy intake. Willett et al. (1) discussed this in their seminal article, and the methods have been further discussed on theoretical and practical levels in several recent articles (2–4). Although intuition and practical subject matter knowledge are important, no formal simulation studies had been performed to date to investigate the implications of different food substitution modeling approaches. In this issue of *The American Journal of Clinical Nutrition*, Tomova et al. (5) fill this gap by investigating several modeling approaches for different food substitutions using simulations with parameters informed by the UK National Diet and Nutrition Survey. Of note, the authors investigate substitutions of 1 food instead of another as well as more complex substitutions of 1 food instead of several others. They also investigate the implications of modeling intake as nutrients, or as foods, either in calorie units or in mass units. Some of their findings confirm previous intuitions: we can rest assured. Others may require researchers to revisit how these models are applied in practice: a cause for concern.

### Rest assured

Food intake at a given time point can be conceptualized as a compositional exposure. Compositional data were defined by Arnold et al. (6) to “(...) comprise the parts of some whole, for which all parts sum to that whole.” Food substitution models take advantage of this feature to, more formally, estimate the relative causal effect of 1 food instead of another. In practice, there are several ways of specifying a food substitution, using either the leave-one-out method or the partition method (3). Tomova et al. in addition define an all-components method (4), which includes all the components that make up the whole. Although including all components is more straightforward when working with nutrients, because there tends to be fewer components, this can also be done with foods or food groups. In their study,

simulations of single food substitutions showed that both the leave-one-out and the partition method could recover an estimate close to the true simulated relative causal effect, with 1 important caveat: given that the same unit was used throughout.

### Cause for concern

When using the leave-one-out method to specify the substitution of meat instead of fish, a commonly used model might include the investigated foods in grams and total energy in calories. The results from the simulation showed that mixing units of measurement reversed the sign of the regression coefficient, whereas models with the same units returned estimates close to the true simulated effect. Given that mixed-units models are often seen in nutritional epidemiology, these results should be a cause for concern. To examine this, we revisit a previous article in which we used such models to investigate substitutions of fish, poultry, or unprocessed red meat instead of processed red meat and risk of type 2 diabetes, adjusted for total energy intake (7). In our study, replacing 150 g/wk of processed red meat with fish, poultry, or unprocessed meat was associated with a slightly lower risk of type 2 diabetes using the leave-one-out method. Revisiting these analyses using the leave-one-out method, estimates were in different directions for mixed-unit (foods in grams and total energy in calories) and same-unit analyses (all in calories), but only for fish or poultry, not unprocessed red meat, instead of processed red meat (Table 1), thus confirming some, but not all, of our previous results. When we used the comprehensive all-components method, either per 150 g/wk or 100 kcal/d, we found similar estimates for each, and similar to the original mixed-unit analysis. Although we found similar associations in our reanalysis, it is important to note that the interpretation of the model with mixed units is, as Tomova et al. (5) call it, obscure.

---

Author disclosures: The authors report no conflicts of interest  
DBI is supported by the Independent Research Fund Denmark, grant 1057-00016B. CCD is supported by Aarhus University.  
Address correspondence to DBI (e-mail: [dbi@ph.au.dk](mailto:dbi@ph.au.dk)).  
Received August 3, 2022. Accepted for publication August 11, 2022.  
First published online October 13, 2022; doi: <https://doi.org/10.1093/ajcn/nqac222>.

**TABLE 1** Food substitution models estimating the relative causal effect of fish, poultry, or unprocessed red meat instead of processed red meat in the Danish Diet, Cancer and Health cohort<sup>1</sup>

Estimand	Model		Estimate, HR (95% CI)
	Name	Formula	
Relative causal effect of fish instead of processed red meat	Mixed unit (150 g/wk), leave-one-out	$T2D = a_0 + a_1 \text{ URM} + a_2 \text{ poultry} + a_3 \text{ fish} + a_4 \text{ total meat} + a_5 \text{ TE} + \text{other}^2 \text{ foods} + \text{covariates}^3$ R1.1 = $a_3$	0.97 (0.94, 1.00) <sup>4</sup>
	Same unit (100 kcal/d), leave-one-out	$T2D = b_0 + b_1 \text{ URM} + b_2 \text{ poultry} + b_3 \text{ fish} + b_4 \text{ total meat} + b_5 \text{ TE} + \text{other foods} + \text{covariates}$ R1.2 = $b_3$	1.02 (0.97, 1.08)
	Same unit (150 g/wk), comprehensive all-components	$T2D = c_0 + c_1 \text{ PRM} + c_2 \text{ URM} + c_3 \text{ poultry} + c_4 \text{ fish} + \text{all other food groups}^5 + \text{covariates}$ R1.3 = $c_4 - c_1$	0.97 (0.94, 1.00)
	Same unit (100 kcal/d), comprehensive all-components	$T2D = d_0 + d_1 \text{ PRM} + d_2 \text{ URM} + d_3 \text{ poultry} + d_4 \text{ fish} + \text{all other food groups} + \text{covariates}$ R1.4 = $d_4 - d_1$	0.96 (0.90, 1.02)
	Mixed unit (150 g/wk), leave-one-out	$T2D = e_0 + e_1 \text{ URM} + e_2 \text{ poultry} + e_3 \text{ fish} + e_4 \text{ total meat} + e_5 \text{ TE} + \text{other foods} + \text{covariates}$ R2.1 = $e_2$	0.96 (0.93, 1.00) <sup>4</sup>
	Same unit (100 kcal/d), leave-one-out	$T2D = f_0 + f_1 \text{ URM} + f_2 \text{ poultry} + f_3 \text{ fish} + f_4 \text{ total meat} + f_5 \text{ TE} + \text{other foods} + \text{covariates}$ R2.2 = $f_2$	1.03 (0.96, 1.10)
Relative causal effect of poultry instead of processed red meat	Same unit (150 g/wk), comprehensive all-components	$T2D = g_0 + g_1 \text{ PRM} + g_2 \text{ URM} + g_3 \text{ poultry} + g_4 \text{ fish} + \text{all other food groups} + \text{covariates}$ R2.3 = $g_3 - g_1$	0.97 (0.94, 1.00)
	Same unit (100 kcal/d), comprehensive all-components	$T2D = h_0 + h_1 \text{ PRM} + h_2 \text{ URM} + h_3 \text{ poultry} + h_4 \text{ fish} + \text{all other food groups} + \text{covariates}$ R2.4 = $h_3 - h_1$	0.97 (0.90, 1.04)
	Mixed unit (150 g/wk), leave-one-out	$T2D = i_0 + i_1 \text{ URM} + i_2 \text{ poultry} + i_3 \text{ fish} + i_4 \text{ total meat} + i_5 \text{ TE} + \text{other foods} + \text{covariates}$ R3.1 = $i_1$	0.97 (0.94, 1.00) <sup>4</sup>
	Same unit (100 kcal/d), leave-one-out	$T2D = j_0 + j_1 \text{ URM} + j_2 \text{ poultry} + j_3 \text{ fish} + j_4 \text{ total meat} + j_5 \text{ TE} + \text{other foods} + \text{covariates}$ R3.2 = $j_1$	0.97 (0.92, 1.02)

(Continued)

TABLE 1 (Continued)

Estimand		Model		Estimate, HR (95% CI)
Name	Formula			
Same unit (150 g/wk), comprehensive all-components	T2D = $k_0 + k_1 \text{ PRM} + k_2 \text{ URM} + k_3 \text{ poultry} + k_4 \text{ fish} + \text{all other food groups} + \text{covariates}$	0.97 (0.94, 1.00)		
	R3.3 = $k_2 - k_1$			
Same unit (100 kcal/d), comprehensive all-components	T2D = $l_0 + l_1 \text{ PRM} + l_2 \text{ URM} + l_3 \text{ poultry} + l_4 \text{ fish} + \text{all other food groups} + \text{covariates}$	0.97 (0.92, 1.02)		
	R3.4 = $l_2 - l_1$			

<sup>1</sup> $n = 53,135$ ; incident cases,  $n = 6877$ ; median follow-up time: 15.4 y. PRM, processed red meat; R, result; TE, total energy; T2D, type 2 diabetes; UPM, unprocessed red meat.

<sup>2</sup>Other food groups: whole grains, fruits, vegetables, dairy products, potatoes, potato chips, soft drinks.

<sup>3</sup>Covariates: BMI (in  $\text{kg/m}^2$ ; restricted cubic spline with 3 knots), waist circumference adjusted for BMI (continuous), smoking (never, former, current), alcohol (g/d; restricted cubic spline with 4 knots), education (<7 y, 8–10 y, >10 y), physical activity ( $\leq$  or >3.5 h/wk), sex (women or men, as strata), baseline date at study entry (tertiles, as strata), and age at baseline (tertiles, as strata).

<sup>4</sup>Results from Ibsen et al. (7), although with slight differences in sample size and cases due to more variables being included.

<sup>5</sup>All other food groups: other foods (legumes + refined cereals + eggs + nuts + vegetable oils + margarines + other animal fats + sweets + mayonnaise + soya + snacks), other drinks (fruit juices + vegetable juices + coffee + tea + water).

Another lesson from the simulation studies by Tomova et al. (5) is the complications that occur when specifying substitutions of 1 food with several others. This situation can be introduced “inadvertently” in models where total energy intake is adjusted for, but no substitution food or nutrient is specified. Again, reversal of coefficient signs occurred when mixing units. Furthermore, the all-components method could only estimate the average relative causal effect of meat (i.e., the effect of meat instead of a weighted average of all other foods) and none of the methods could sufficiently recover the relative causal effect of meat instead of cereal, dairy, fish, nuts, and other foods in grams.

### The continued evolution of food substitution methods

A key message from Tomova et al. (5) is that nutrition researchers ought to think carefully about the formulation of their research questions, the estimates of interest, and how the chosen statistical models reflect these. One way of doing this is to treat food intake as a compositional exposure and make this explicit in the chosen statistical model, which could be emphasized by using the all-components method. The article by Tomova et al. (5) encourages us to revisit our thinking of food substitution models, whether inadvertent or not. Future developments will want to focus on dietary changes and here the target trial framework (8) with g-methods (9, 10) to specify dietary interventions may aid in specifying food substitutions over multiple time points.

Nutrition epidemiology is often criticized for not providing valid results, but dietary guidelines and nutrition policies rely on such evidence for diseases of long latency and public health burden. We believe that a key reason for this criticism, which may stem from conflicting results from subsequent studies, is that the research questions and chosen methods in seemingly similar studies are often not precise enough to allow comparison. Given the complexity of diet as an exposure, there certainly is a need for further methodological developments to improve the methods to

handle such complex questions. This article by Tomova et al. (5) is 1 such example.

The authors' responsibilities were as follows—DBI: conducted the analyses; and both authors: wrote the first draft, edited the manuscript, and read and approved the final manuscript.

### Data Availability

Data described in this article may be made available upon request pending on application to and approval by the Danish Cancer Society (e-mail: dchdata@cancer.dk)

### References

1. Willett WC, Howe GR, Kushi L. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997;65(4):1220S–8S.
2. Song M, Giovannucci E. Substitution analysis in nutritional epidemiology: proceed with caution. *Eur J Epidemiol* 2018;33(2):137–40.
3. Ibsen DB, Laursen ASD, Würtz AML, Dahm CC, Rimm EB, Parner ET, et al. Food substitution models for nutritional epidemiology. *Am J Clin Nutr* 2021;113(2):294–303.
4. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. *Am J Clin Nutr* 2022;115(1):189–98.
5. Tomova GD, Gilthorpe MS, Tennant PWG. Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. *Am J Clin Nutr* 2022;116(5):1379–88.
6. Arnold KF, Berrie L, Tennant PWG, Gilthorpe MS. A causal inference perspective on the analysis of compositional data. *Int J Epidemiol* 2020;49(4):1307–13.
7. Ibsen DB, Warberg CK, Würtz AML, Overvad K, Dahm CC. Substitution of red meat with poultry or fish and risk of type 2 diabetes: a Danish cohort study. *Eur J Nutr* 2019;58(7):2705–12.
8. Hernan MA. Methods of public health research—strengthening causal inference from observational data. *N Engl J Med* 2021;385(15):1345–8.
9. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model* 1986;7(9–12):1393–512.
10. Chiu Y, Chavarro JE, Dickerman BA, Manson JE, Mukamal KJ. Estimating the effect of nutritional interventions using observational data: the American Heart Association's 2020 Dietary Goals and mortality. *Am J Clin Nutr* 2021;114(2):690–703.